

On the Impact of Layout Quality to Understanding of UML Diagrams: Not Just Pretty Pictures

Harald Störrle

Department of Applied Mathematics and Computer Science

Technical University of Denmark

Matematiktorvet, 2800 Lyngby, Denmark

hsto@dtu.dk

Abstract: In a string of empirical studies, we show that the layout of UML diagrams contributes significantly to understanding the underlying model. This effect extends over different factors such as diagram type & size, and expertise.

Status Quo The Unified Modeling Language (UML) has been called the “*lingua franca of software engineering*” for over a decade now. It is widely believed that, to a sizable degree, its popularity is rooted in the predominantly visual nature of UML models. The advantage of visual over textual notations is, generally speaking, that they support human perceptual and thought processes, making diagrams a more cognitively efficient medium than, say, prose.

Practical experience suggests that the usage and understanding of UML diagrams is greatly affected by the quality of their layout. While existing research failed to provide conclusive evidence in support of this hypothesis, our own work [Stö11, Stö12, Stö14] provided substantial evidence to this effect.

Size Matters When analyzing the impact factors, we find that diagram size is an important factor to diagram understanding; this is consistent with previous findings [MRC07]. Other factors like expertise level are important, too, though to a lesser degree, and some factors appear to be irrelevant, such as diagram type. Since there was no adequate definition of this notion, we had to define diagram size metrics first. It turns out that the most trivial notion of simply counting diagrammatic elements is highly correlated to more complex notions, an effect known from program size metrics. By Occams razor, thus, we conclude that the size of a diagram should be measured as the number of diagram elements (i.e., geometric shapes, annotations, and line segments).

Studying the impact of diagram size to diagram understanding by modelers, we find that there is a strong negative correlation between size and performance as well as preference. Our results are statistically highly significant and far exceed earlier work in terms of validity in several dimensions.

We utilize these results to derive recommendations on diagram sizes that are optimal for model understanding. More recent work has begun to uncover the cognitive mechanisms involved in the understanding of UML diagrams [Mai14, SBCM14].

Generalizations It is a common misunderstanding is to confuse or identify “model” with “diagram”: the *model* is the underlying data structure while the *diagram* is its visual representation. Our work refers exclusively to diagrams. From a visual languages point of view, there are only minor differences between the notations of many of the popular modeling languages, and diagram type does not seem to be an important factor of modeler understanding, anyway. Thus, our results are probably generalizable to BPMN, EPCs and so on. The substantial differences between such modeling languages on the conceptual, semantic, and pragmatic levels are largely irrelevant for the aspects studied in our work.

Next Steps We are currently working on two threads to continue this research. First, we are in the final stages of a systematic literature review on the field of UML diagram layout knowledge, so as to complement and frame our own findings in a scientifically reliable way. Second, we would like to increase the validity of our findings through independent replications. To this end, we are creating an online test implementing the experimental setup used in our work. It will allow anybody to administer, or individually participate in our experiment. Over time, we hope to acquire data from a large enough population so that we can find general averages. Thus, effectively, our experiment might evolve into a personal performance inventory not unlike existing IQ tests. This could be used as a diagnostic instrument, e.g., for self-assessment and as an experimental pre-testw.

References

- [Mai14] Maier, Anja M. and Baltsen, Nick and Christoffersen, Henrik and Störrle, Harald. Towards Diagram Understanding: A Pilot-Study Measuring Cognitive Workload Through Eye-Tracking. In *Proc. Intl. Conf. Human Behavior in Design*, 2014.
- [MRC07] Jan Mendling, Hajo A. Reijers, and Jorge Cardoso. What Makes Process Models Understandable? In G. Alonso, Peter Dadam, and Michael Rosemann, editors, *Proc. Intl. Conf. Business Process Management (BPM)*, pages 48–63. Springer Verlag, 2007.
- [SBCM14] Harald Störrle, Nick Baltsen, Henrik Christoffersen, and Anja M. Maier. On the Impact of Diagram Layout: How Are Models Actually Read? In S. Sauer, M. Wimmer, M. Genero, and S. Qadeer, editors, *Joint Proc. MODELS 2014 Poster Session and ACM Student Research Competition*, volume 1258, pages 31–35. CEUR, 2014.
- [Stö11] Harald Störrle. On the Impact of Layout Quality to Understanding UML Diagrams. In *Proc. IEEE Symp. Visual Languages and Human-Centric Computing (VL/HCC’11)*, pages 135–142. IEEE Computer Society, 2011.
- [Stö12] Harald Störrle. On the Impact of Layout Quality to Understanding UML Diagrams: Diagram Type and Expertise. In G. Costagliola, A. Ko, A. Cypher, J. Nichols, C. Scaffidi, C. Kelleher, and B. Myers, editors, *Proc. IEEE Symp. Visual Languages and Human-Centric Computing (VL/HCC’12)*, pages 195–202. IEEE Computer Society, 2012.
- [Stö14] Harald Störrle. On the Impact of Layout Quality to Understanding UML Diagrams: Size Matters. In Jürgen Dingel, Wolfram Schulte, Isidro Ramos, and Emilio Abrahao, Silviaand Insfran, editors, *Proc. 17th Intl. Conf. Model Driven Engineering Languages and Systems (MoDELS’14)*, number 8767 in LNCS, pages 518–534. Springer Verlag, 2014.